

Overreporting: ein vernachlässigtes Problem und die Schwierigkeit von Gegenmaßnahmen

Hardmeier, Sibylle; Fontana, Marie-Christine

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Hardmeier, S., & Fontana, M.-C. (2006). Overreporting: ein vernachlässigtes Problem und die Schwierigkeit von Gegenmaßnahmen. *ZUMA Nachrichten*, 30(58), 50-80. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-211132>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

OVERREPORTING: EIN VERNACHLÄSSIGTES PROBLEM UND DIE SCHWIERIGKEIT VON GEGENMASSNAHMEN

OVERREPORTING: A NEGLECTED ISSUE IN GERMAN-LANGUAGE LITERATURE AND THE PROBLEM OF COUNTERMEASURES

SIBYLLE HARDMEIER & MARIE-CHRISTINE FONTANA

Overreporting wird in der deutschsprachigen Methodenliteratur kaum behandelt – zu Unrecht, wie ein Blick auf die empirischen Daten zeigt. Der vorliegende Artikel hat sich daher zum Ziel gesetzt, dieses in Vergessenheit geratene Problem in Erinnerung zu rufen, und geht auf drei Ebenen vor: Im ersten Teil präsentieren wir einen Überblick über den beachtlichen Literaturkorpus im englischsprachigen Raum und nehmen eine Systematisierung entlang methodischer Problemstellungen und theoretischer Begründungen des Phänomens vor – der sozialen Erwünschtheit einerseits und dem Misremembering andererseits, das als Erinnerungsproblem oder als „source confusing“ verstanden werden kann. Im zweiten Teil widmen wir uns den zwei theoretischen Erklärungssets anhand empirischer Daten aus der Schweiz. Mit der Hilfe von Aggregatdatenanalysen sowie dreier Feldexperimente nehmen wir erste explorative Hypothesentests vor und analysieren den Effekt theoretisch hergeleiteter Maßnahmen auf der Ebene des Fragebogens (Wording). Es zeigt sich, dass Gegenmaßnahmen zur sozialen Erwünschtheit mit kleinen Modifikationen im Wording schwierig zu erreichen sind und sogar kontraproduktiv sein können. Ein Effekt des Treatments zum Misremembering ist bei großen Zeitabständen zwar vorhanden, jedoch bleibt die kausale Begründung offen und schwierig. Daher schlagen wir im letzten Teil weitere Maßnahmen für künftige Untersuchungen vor: Ein Bogus-pipeline-Experiment, Kontrollen über Erwünschtheits- bzw. Erinnerungsskalen, eine Meta-Analyse und Validitätsstudien für jene Länder, die noch keine haben.

Overreporting is hardly discussed in German-language methodology literature – hardly justifiable, as a glance at the empirical data shows. The intention of the present article is therefore to bring this forgotten problem back to memory. We proceed at three levels: the first part presents an overview of the considerable body of literature

found in English-speaking countries and classifies it according to methodological approaches and theoretical explanations of the phenomenon – social desirability on the one hand and misremembering on the other, which can be interpreted as a memory failure or as “source confusing”. The second part is devoted to two theoretical sets of explanations, using empirical data from Switzerland. We conduct some initial explorative hypotheses tests based on aggregate data analyses and three field experiments, and examine the effect of theoretically deduced measures at the level of questionnaires (wording). This reveals that it is difficult to create countermeasures to social desirability with minor modifications of the wording and that it may even be counter-productive. Despite treatment of misremembering having an effect when time spans are sufficiently large, the causal explanation remains unclear and difficult. The last part therefore proposes additional measures for future investigations: a bogus pipeline experiment, checks on desirability and memory scales, a meta-analysis and validity studies for countries still lacking them.

1 Einleitung

In einer Zeit, in der wir beinahe täglich mit den Ergebnissen von Befragungsdaten konfrontiert werden, geht leicht vergessen, dass die Demoskopie ein sensibles Erhebungsverfahren darstellt und mit Selektions- ebenso wie mit Messproblemen (Groves 1989) konfrontiert ist. Am Beispiel des so genannten Overreporting – der Tatsache also, dass bei Wahlbefragungen der Anteil jener, die angeben, an der Wahl teilgenommen zu haben, höher ist als die tatsächliche Wahlbeteiligung – ist das besonders dokumentiert. Die höhere, in den Befragungsdaten angezeigte Partizipationsrate wird nämlich, abgesehen von Referenzproblemen in der offiziellen Statistik¹ und allfälligen Stimuluseffekten in Panels, im Wesentlichen auf zwei unterschiedliche, allenfalls jedoch gleichzeitig auftretende Ursachen zurückgeführt:

- dass Nichtwählende in der gezogenen Stichprobe untervertreten sind (Sampling-Fehler);
- und dass einige Befragte eine Fehlauskunft zu ihrem vermeintlichen Teilnahmeverhalten geben, womit das Overreporting zu einem speziellen Fall von Misreporting wird (Messfehler).

1 Die offizielle Statistik kann zu viele Wahlberechtigte ausweisen und damit den Nichtwähleranteil überschätzen, wenn sie Wegzüge, Todesfälle oder Ähnliches nicht auszuweisen vermag. Vgl. Clausen (1968); Kleinhenz (1995); Lavies (1973); Presser et al. (1990).

Bereits bei der Durchführung der ersten repräsentativen Befragungen in den USA wurde die politische Demoskopie mit diesem ernsthaften Validitätsproblem² konfrontiert. Wie eine Recherche in JSTOR zeigt,³ beschäftigt sich die amerikanische Zunft auch heute noch mit diesem Problem. Im Gegensatz dazu scheint die Auseinandersetzung – obwohl auch im Kontext europäischer Wahlen früh konstatiert⁴ – im deutschsprachigen Raum deutlich weniger intensiv. Eine entlang identischer Suchkriterien⁵ wie bei JSTOR durchgeführte Literaturrecherche in den einschlägigen wissenschaftlichen deutschsprachigen Fachzeitschriften⁶ generiert keinen einzigen Hit.⁷

Ist dieses Schweigen gerechtfertigt und sind die Mahnrufe von frühen Protagonisten der deutschen (Kaase 1973) oder schweizerischen Wahlforschung (Steiner 1965) heute obsolet? Oder müssen wir vielmehr eine wichtige Forschungslücke benennen? Ein flüchtiger Blick in die schweizerischen und deutschen Politdaten gibt eine erste Antwort auf diese Fragen: Im Politbarometer 2002 gaben 89.8 Prozent der Befragten im September 2002 an, sie würden sicher zur Wahl gehen oder hätten bereits brieflich gestimmt. Weitere 6.4 Prozent hielten fest, sie würden wahrscheinlich teilnehmen. Gemessen an der faktischen Beteiligung vom 22. September 2002 von 79.1 Prozent ist damit von einem Overreporting von 10.7 Prozentpunkten auszugehen.⁸ Gleichtags wurden die Schweizer Stimmberechtigten zu einer Referendumsabstimmung an die Urnen gerufen. In der VOX-Nachbefragung gaben 61.2 Prozent der Befragten an, sie hätten ihre Stimme abgegeben; die effektive Beteiligung lag jedoch nur bei 45.2 Prozent, womit sich ebenfalls ein deutliches Overreporting abzeichnete (15 Prozentpunkte).

2 In den meisten Overreporting-Studien wird aufgrund der Annahme von systematischen Messfehlern das Problem der Validität in der Tat als bedeutender angesehen als jenes der Reliabilität (Carmines & Zeller 1979: 13 ff). Kritisch dazu äußern sich Andersson & Granberg (1997: 127).

3 Eine Recherche mit dem Suchbegriff „Overreporting“ in den einschlägigen Journals der Sozialwissenschaft hat nicht weniger als 64 Zeitschriftenartikel zu Tage gefördert; wobei 14 davon als eigentliche Overreporting-Studien zu taxieren sind.

4 Swaddle & Heath 1989; Kleinhenz 1995; Rattinger & Wagner 2003; van der Kolk & Aarts 2001, Granberg & Holmberg 1991 sowie Riklin & Kley 1981.

5 Also: <over> sowie <antw>, <antwortstil>, <messfehl>, <erinnerung>, <response>.

6 Kölner Zeitschrift für Soziologie und Sozialpsychologie (Gesamtinhaltsverzeichnis der Titel 2003-1996); Politische Vierteljahresschrift (Jahresinhaltsverzeichnis der Titel 2003-2000); ZA-Informationen (Integrale Ausgaben ab Nr. 37 sowie Titel und Abstracts der Nummern 1-36); ZUMA-Nachrichten (Gesamtverzeichnis der Titel 2003-1977) sowie ZUMA-Arbeitsberichte (Gesamtverzeichnis der Titel 2003-1980).

7 Dabei kann das nicht mit einer Abneigung gegenüber anglistischen Termini begründet werden: Zum einen finden andere Anglizismen wie „Survey“, „Bias“ oder „Response“ durchaus Anklang. Zum anderen wird der Fachterminus „Overreporting“ in Buchpublikationen durchaus verwendet (Kaase & Bauer-Kaase 1998: 87). Allerdings gilt es festzuhalten, dass eines der wichtigsten Survey-Methodenbücher aus Deutschland (Schnell 1997) sich ‚nur‘ mit dem Unit Response beschäftigt und den Ausfall von Nichtwählern nicht speziell untersucht.

8 Politbarometer West 2002 ZA-Nr. 3849, S. 25 sowie Statistisches Bundesamt, Bundestagswahl 2002, Heft 3, S. 6.

Der vorliegende Artikel widmet sich daher dem Phänomen des Overreporting – nachfolgend mit OR abgekürzt – und verfolgt folgende Ziele: Wir möchten erstens den aktuellen Forschungsstand aufarbeiten und dem interessierten Publikum zugänglich machen. Die beachtliche Anzahl Publikationen aus dem englischsprachigen Raum werden in Kapitel 2 anhand theoretischer und methodischer Kriterien systematisch diskutiert. Zudem werden die Resultate einiger empirischer Primär- und Sekundärdatenerhebungen präsentiert und im aktuellen Forschungsstand verortet, wobei der Effekt des Wordings im Vordergrund steht. Daten und Experimente aus der Schweiz, die in Kapitel 3 präsentiert werden, bestätigen den Eindruck, dass Gegenmaßnahmen in diesem Bereich schwierig und zwiespältig sind. Vor diesem Hintergrund ziehen wir in Kapitel 4 einige Folgerungen und präsentieren Anregungen für weitergehende Untersuchungen.

2 Forschungsstand

In den USA ist die wissenschaftliche Auseinandersetzung mit dem OR so alt wie die Wahlforschung. Nachdem bei den Wahlgängen von 1936 und 1948 die Wahlprognosen kräftig daneben lagen, begann die Disziplin, Validitätsprobleme von Wahlbefragungen theoretisch zu erörtern, das Phänomen des OR deskriptiv zu erfassen und empirische Validitätsstudien durchzuführen. Diese dienten dazu, Profile der Misreporter zu erstellen und kausale Erklärungen zu präsentieren. Gleichzeitig begann man mögliche Korrekturmechanismen zu entwickeln und die allfällige Verzerrung von Modellschätzungen zu erörtern.

2.1 Overreporting als Misreporting

Bereits die ersten deskriptiven Untersuchungen zum Phänomen des OR konnten einen recht konsistenten Befund festhalten: Sobald es nicht um statische Faktbeschreibungen wie den Besitz eines Telefons, sondern um „ex-post“ Angaben zu einem Verhalten geht, das normativ aufgeladen ist, sind die Befragungsdaten in eine Richtung verzerrt. Auch bei den amerikanischen Wahlstudien wurde dieses Phänomen von Beginn an beobachtet und die Abweichungen auf aggregiertem Niveau bewegen sich zwischen 13 und 15 Prozentpunkten (Presser et al. 1990). Allerdings ist aufgrund des Vergleichs ‚Befragungsdaten – offizielle Statistik‘ das Phänomen des Overreporting unterbeleuchtet. Daher wurden in den USA schon früh so genannte Validitätsstudien realisiert, in welchen die selbst bekundete Partizipation mit offiziellen Daten aus den Wahlbüros validiert wurde. Der Vorteil dieses Zugangs liegt auf der Hand: Erst diese Analysen konnten bestätigen, dass das auf aggregiertem Niveau beobachtete Phänomen des OR in vielen Fällen ein Messproblem darstellt und mit einem Misreporting auf individueller Ebene, also bei den Befragten, einhergeht.

Dabei machten erste und thematisch breit angelegte Studien vorerst auf die Verbreitung des Phänomens bei so genannten self-reports in Befragungen aufmerksam. Misreporting findet bei verschiedenen Themen (Cahalan 1968; Parry & Crossley 1950) wie auch unterschiedlichen Partizipationsformen (Volgy & Schwarz 1984) statt und wird offensichtlich durch eine gemeinsame Klammer zusammengehalten: Es tritt auf, sobald die Themen normativ aufgeladen und die entsprechenden Fragen als sensitiv wahrgenommen werden (Renzetti & Lee 1993). Daher kann auch eine Grundregel festgehalten werden (Tourangeau et al. 2000: 269 ff.; Belli et al. 2001; Sigelman 1982): Befragte neigen zu einem Underreporting bei sozial unerwünschten und zu einem Overreporting bei sozial erwünschten Themen. Entsprechend sind Fragen zur politischen Partizipation nicht nur besonders anfällig für diese Verzerrung (Sudman & Bradburn 1974: 59), sondern die Richtung des Misreporting entpuppt sich – von einigen wenigen Ausnahmen abgesehen⁹ – als „Einbahnstraße“ (Eilfort 1994: 90): Das Misreporting ist ein Overreporting.

Allerdings blieben trotz einer frühen Mahnung von Anderson und Silver (1986: 775) die Berechnungsweisen des OR in diesen Validitätsstudien sehr heterogen und zuweilen wenig überzeugend. Denn oft wird für die Berechnung der Anteile nicht der richtige Nenner gewählt und die eigentliche „population at risk“ gerät aus dem Blickfeld:¹⁰ jene Nichtwählenden nämlich, welche unter Umständen sozial stigmatisiert werden, wenn sie nicht teilnehmen. Wenn man also der Empfehlung von Anderson und Silver folgt und misst, wie viele von den validierten Nichtwählenden behaupten, gewählt zu haben, dann ist das OR beträchtlich. So liegt der Wert der wohl umfassendsten Validitätsanalyse für die ANES zwischen 1964 und 1988 nicht wie berichtet bei rund 14 Prozent (Traugott 1989: 8), sondern bei rund 25 Prozent. Bei Swaddle & Heath käme man entsprechend auf 24.9 Prozent und würde sich wiederum auf ähnlich hohem Niveau wie bei Sigelman (1982: 49), Andersson & Granberg (1997), Granberg & Holmberg (1991) oder Karp & Brockington (2005 forthcoming) bewegen. Das Misreporting scheint also nicht nur ein pankulturelles Phänomen darzustellen, sondern bewegt sich auch auf hohem Niveau.

9 Auf wichtige Ausnahmen und damit systematisches Underreporting bei der Teilnahmefrage – wenn „low-salience“ Wahlen zeitgleich zu „high-salience“ Wahlen stattfinden – haben vor allem Adamany & Shelley (1980) aufmerksam gemacht. Jennings hat in seiner nach Staaten differenzierten Studie festgestellt, dass in Alaska konstant ein Underreporting stattfindet (siehe Presser et al. 1990: 3).

10 Einige Autoren definieren nämlich Misreporting als den Gesamtanteil unkorrekter Antworten (Steiner 1965; Volgy & Schwarz 1984; Cahalan 1968). Andere Autorinnen wie zum Beispiel Sigelman (1982) oder Traugott & Katosh (1979) referieren den Anteil unter jenen, die eine Teilnahme behauptet haben, tatsächlich aber der Urne fernblieben.

2.2 Das Profil der Overreporter

Wenn Validitätsstudien neben der Teilnahmefrage zusätzliche soziodemographische Variablen erfassen, dann können sie einen Beitrag zur Profilbildung der Overreporter liefern. Ein Blick auf die Literatur zeigt allerdings, dass die Erstellung eines solchen Sozialprofils trotz einer Vielzahl realisierter Studien nur beschränkt möglich ist. Neben den Zugangsweisen¹¹ sind auch die Befunde disparat. Während beispielsweise die einen Schwarze oder Nichtweiße eher als Overreporter identifizieren,¹² wird dies von Anderson & Silver (1986) methodisch in Frage gestellt und von Bernstein et al. (2001) differenziert. Da zudem überraschend oft die jüngeren Stimmberechtigten – von denen man vermuten könnte, sie hätten die soziale Norm der Teilnahme weniger verinnerlicht – zu den Overreportern zählen,¹³ trifft die Einschätzung, dass die Overreporter eher den Wählenden als den Nichtwählenden ähneln, nur beschränkt zu. In der Tendenz – und vor dem Hintergrund von methodisch sophistizierteren Analysen (Anderson & Silver 1986; Silver et al. 1986; Bernstein et al. 2001) – lässt sich aber die These aufstellen, dass es vor allem die politisch und gesellschaftlich Involvierten sind, welche zum OR neigen.

2.3 Theoretische Begründungen: Soziale Erwünschtheit und Misremembering

Die regelmäßig durchgeführten Validitätsstudien der ANES haben den Eindruck erhärtet, dass es sich beim OR der Wahlteilnahme um ein Messproblem handelt. Als wichtigste Erklärung dafür wurde die soziale Erwünschtheit aufgeführt. Dieser formale Antwortstil (Crowne & Marlowe 1964) beschreibt „the tendency of people to deny socially undesirable traits and to ‘admit’ socially desirable ones” (Phillips 1973: 20). Das heißt mit anderen Worten, der Begriff des “Lügens” vermag zwar das Antwortverhalten der Befragten faktisch zu umschreiben, hinsichtlich der dahinter liegenden Motivation ist er aber wenig erhellend (Mummendey 1999: 160). Die spezifische Strategie der Selbstpräsentation in Form von OR hat vielmehr mit sozialen Normen und dem Bedürfnis nach sozialer Anerkennung zu tun (Scholl 1993: 67 ff.); wobei diese Strategie der „intentional deception“ (Belli et al. 2001: 494) manchmal eher als Persönlichkeitszug, dann wiederum eher als Situationsreaktion verstanden werden kann.¹⁴

11 Dabei ist vor allem auf zwei Differenzen hinzuweisen: Weil erstens, wie in Abschnitt 2.1 ausgeführt, unterschiedliche Populationen bzw. Nenner betrachtet werden, basieren die Sozialprofile auf unterschiedlichen Gruppenvergleichen. Zweitens ist das analytische Vorgehen unterschiedlich. Während sich die einen bei der Profilerstellung auf deskriptive Zugänge beschränken, versuchen die anderen, das Profil der Overreporter kausal zu erklären (Traugott & Katosh 1979; Katosh & Traugott 1981).

12 Abramson & Claggett (1986); Belli et al. (2001); Hill & Hurley (1984); Sigelman (1982).

13 Bernstein et al. 2001; Cahalan 1968; Traugott & Katosh 1979; Wolfinger & Rosenstone 1980.

14 Eine entsprechende Unterteilung der Motive findet sich auch in den stärker von der Rational-choice-Theorie inspirierten Ansätzen der Wert-Erwartungstheorie (Esser 1991; Tourangeau et al. 2000).

Allerdings entpuppt sich vor diesem Hintergrund das response set „Overreporting“ auch als ein recht volatiles und kontextabhängiges Konstrukt, das Prognosen darüber, wer zu den Overreportern gehört, nicht einfach macht. Insgesamt ist sich aber die Literatur – auch vor dem Hintergrund der Ausführungen unter Punkt 2.3 – einig, dass es eher die Personen mit höherer Bildung und stärkerem Involvierungsgrad sind. Diese Personengruppen seien sich der gesellschaftlich ‚korrekten‘ Antworten stärker bewusst (Silver et al. 1986: 623). Die Beteiligung an Wahlen und Abstimmungen ist eben nicht nur ein freier Entscheid, sondern stellt auch eine sozial vermittelte Bürgerpflicht dar. OR aus Gründen der sozialen Erwünschtheit setzt also ein entsprechendes Pflichtbewusstsein voraus. Stark verknüpft, wenn auch nicht immer expliziert, mit der sozialen ErwünschtheitsThese sind Vermutungen über den Einfluss der Erhebungsmethode auf das OR sowie über den Wandel des OR im Laufe der Zeit:

Allgemein wird angenommen, dass die Erhebungsmethode mit dem höheren Grad an Anonymität und weniger direktem Kontakt mit Interviewern – das Telefoninterview – weniger soziale Erwünschtheit allgemein und damit auch weniger OR provoziert. Während Untersuchungen zu sensitiven Fragen diese These bestätigen (de Leeuw & van der Zouwen 1988; Tourangeau et al. 2000: 295) und weitgehend erhärtet ist, dass anonyme Verfahren die Verzerrung durch soziale Erwünschtheit reduzieren (Richman et al. 1999: 756), ist der Literaturkorpus für das OR noch schmal und liefert disparate Ergebnisse.¹⁵ Einige Studien können zumindest einen schwachen, positiven Effekt der telefonischen Erhebungsmethode nachweisen (Locander et al. 1976; Rogers 1976). Andere können das nicht bestätigen (Abramson & Claggett 1986: 416) bzw. sehen keine Effekte der Erhebungsmethode (Voogt & Saris 2003: 170 f.).¹⁶

Für moderne Demokratien wird häufig ein Prozess der Individualisierung diagnostiziert, der sich in der Wahlforschung in abnehmenden Parteibindungen und Teilnahmeraten niederschlägt. Die Beteiligung an Wahlen und Abstimmungen – so die allgemeine Diagnose – wird stärker als freiwillige Handlung und weniger als soziale Pflicht angesehen. Folglich sollte auch die soziale Erwünschtheit der politischen Teilnahme abgeschwächt worden sein.¹⁷ Allerdings bestätigen die bisher publizierten Daten eher das Gegenteil.

15 Das Argument geht auf Sudman & Bradburn (1974: 9 f.) zurück. Der National Household Survey on Drug Abuse zum Beispiel verwendet deshalb selbst administrierte, schriftliche Befragungen. Zu Ausnahmen beim OR vgl. auch Voogt & Saris (2003: 170). Es stellt sich vor allem die Frage, ob schriftliche oder Online-Erhebungen neben der ‚Anonymität‘ nicht auch einen höheren Grad an Verbindlichkeit vermitteln.

16 Ein Blick auf das Design der Studie (ebd: 168) lässt allerdings vermuten, dass hier Sampling- und Erhebungseffekte zusammenwirken.

17 In einer Querschnittsanalyse bestätigen Hill & Hurley (1984) diese These indirekt. Aufrichtige Nichtwähler haben einen schwächer ausgeprägten Sinn für die Staatsbürgerpflicht.

Während Belli et al. (2001: 494) tatsächlich eine Abnahme im OR konstatieren, gibt es überzeugende, wenn auch nicht ungeteilte Hinweise dafür, dass diese Diskrepanz zwischen offizieller Statistik und Selbstdeklaration in der jüngsten Zeit eher zugenommen hat (Bernstein et al. 2001; Burden 2003; Burden 2000; van der Kolk & Aarts 2001).¹⁸ Eine nahezu konstante Diskrepanz zwischen faktischer und bekundeter Teilnahme geht nach Burden mit einer abnehmenden Wahlbeteiligung und somit einem steigenden OR einher – unabhängig davon, wie man die Wahlbeteiligung misst.

Die kognitive Wende in der Sozialwissenschaft und die Sensibilisierung dafür, dass Prozesse der Informationsverarbeitung ebenso wie die Beantwortung von Survey-Fragen komplexe und mehrstufige Verfahren darstellen, haben in der Befragungsforschung das Interesse für so genannte Erinnerungs- (recall) und Abrufungsprobleme (retrieval) geweckt (Tourangeau et al. 2000: 7 ff.). Das hat auch Einfluss auf die OR-Debatte. Ergänzend zur Intentional-deception-Hypothese wird in jüngster Zeit verstärkt die Misremembering-Hypothese ins Feld geführt (Belli et al. 2001: 494). Mit dem differenzierteren Wissen über Gedächtnisstrukturen und der Sensibilisierung dafür, dass der Prozess des Retrievals durch verschiedene Faktoren gestört werden kann, erscheint das Antworten auf Fragen keineswegs als banale Angelegenheit. Deshalb begann man neben der Frage zum Stimmverhalten auch die Partizipationsfrage vor diesem kognitionspsychologischen Hintergrund zu untersuchen (Abelson et al. 1992; Belli et al. 1999 bzw. 2001).

Dabei können ‚simple‘ Erinnerungsprobleme ausschlaggebend sein – insbesondere dann, wenn der Erinnerungsgegenstand weit zurückliegt. Allerdings ist es zumeist nicht die Zeitspanne allein, welche einen negativen Einfluss hat, sondern die Tatsache, dass in der Zwischenzeit andere oder sogar ähnliche Ereignisse stattgefunden haben. Nicht Erinnerung, sondern Retrievement wird damit die kritische Größe, weil das so genannte „source monitoring“ gestört wird (Belli et al. 2001: 480; Tourangeau et al. 2000: 83). Das bedeutet indessen auch, dass sich das Profil der potenziell Anfälligen ändern sollte: Während einfache Erinnerungsprobleme bei den weniger Gebildeten und politisch weniger Interessierten größer sein sollten (Adamany & Shelley 1980), sind vor allem die Involvierten vom source confusing betroffen. Nur jene, welche die Absicht hatten an der Wahl teilzunehmen, können Intentionen und Handlungen unter Umständen verwechseln; und nur wenn andere politische Partizipationsformen in der Zwischenzeit stattgefunden haben, kann Konfusion entstehen.

18 Burden kann die Kritik (Martinez 2003; McDonald 2003) an seinem Artikel aus dem Jahre 2000, wonach die vermeintliche Zunahme lediglich auf die unterschiedlichen Grundgesamtheiten der Wahlstudien und offiziellen Statistiken zurückzuführen sei, empirisch widerlegen.

Aufgrund dieser Tatsache und im Wissen darum, dass das Misreporting nicht zufällig, sondern gerichtet ist, schlagen Belli und seine Kolleginnen (Belli et al. 2001; Belli et al. 1999) eine Kombination der beiden Erklärungsansätze vor. Die soziale Erwünschtheit spiele vor allem dann eine Rolle, so ihre Annahme, wenn Erinnerungsprobleme auftreten. Andererseits müssten Erinnerungsprobleme allein, ohne Kombination mit sozialer Erwünschtheit, genau so oft zu Under- wie zu Overreporting führen. Da dies bekanntermaßen nicht der Fall ist, sei ein Zusammenspiel der beiden Erklärungsfaktoren plausibel (Belli et al. 1999: 91).

2.4 Einfluss auf Modellschätzungen und Gegenmaßnahmen

Nachdem sich die ANES zunächst prioritär den Profilen und Ursachen des OR gewidmet hatte, provozierte nicht zuletzt das Streichen der Validitätsanalysen im Jahre 1990 einen Wandel in der Forschungsagenda. So erlaubten es die mittlerweile großen Zeitreihen, den oben besprochenen Einfluss von Kontexteffekten, von neuen Erhebungsmethoden oder abnehmenden Parteibindungen und Beteiligungsraten zu erörtern. Darüber hinaus wurde der Frage nach möglichen Effekten auf Modellschätzungen wieder stärkere Aufmerksamkeit zuteil und die Befunde verweisen vermehrt in die Richtung, die bereits Hyman (1945: 559) aufgezeigt hatte: Die Parameter der Wahl- und Abstimmungsforschung sowie Modellschätzer von multivariaten – linearen und nichtlinearen – Modellen können durch das Problem des OR verzerrt werden (Cassel 2003; Bernstein et al. 2001; Voogt 2000; Atkeson 1999).

Insofern ist es nicht erstaunlich, dass die Diskussion über Gegenmaßnahmen zum OR die Forschungsagenda bis heute prägt. Dabei haben in Analogie zur theoretischen Ausrichtung und Begründung bis jetzt Zugänge überwogen, welche beim Messfehler angesetzt haben. In der Terminologie der Befragungsmethodik standen also so genannte Kontexteffekte im Vordergrund. Der Partizipationsfrage vorangestellte Fragen zu früheren oder anderen Teilnahmeforen könnten – so die Vermutung von Presser (1990) – den sozialen Druck abfedern bei der eigentlich interessierenden Frage, sich ins positive Licht zu rücken, bzw. die Verlegenheit über eine gelegentliche Nichtteilnahme zu reduzieren. Ein entsprechender experimenteller Test konnte die Annahme jedoch nicht bestätigen. Auch in den Experimenten von Abelson und seinem Team musste diese These des „two-time-frame procedure“ (Abelson et al. 1992) verworfen werden. Dennoch hatten diese Tests und insbesondere die Schlussfolgerungen große Ausstrahlung. Unverkennbar haben sie Belli und sein Team für ihre weiteren Untersuchungen inspiriert: Nach einem ersten erfolglosen Versuch (Belli et al. 1994) konnten sie denn auch die von Abelson et al. entwickelte Hypothese bestätigen: In zwei telefonischen „split-half“ Experimenten schnitt eine Formulierung, welche darauf abzielte, Probleme des source confusing zu beheben, besser ab als die ANES-Standardfrage und erwies sich besonders bei großen Zeitabständen zum befragten Ereignis als effektiv (Belli et al. 2001).

3 Analysen zum Overreporting im Kontext schweizerischer Wahl- und Abstimmungsstudien

Der schweizerische Kontext stellt ein ideales Forschungsfeld zur Überprüfung von verschiedenen OR-Thesen dar. Zwar liegen keine neueren Validitätsstudien vor, allein der Vergleich zwischen verschiedenen Typen von Urnengängen (Wahlen und Abstimmungen), zwischen Urnengängen mit ganz unterschiedlichen Partizipationsraten bei Sachabstimmungen ebenso wie die Zeitreihe für Sachabstimmungen erlauben es aber, einige Grundhypothesen zum OR zu testen. Darüber hinaus hatten wir die Möglichkeit, bei drei Untersuchungen je einen kleineren Methodentest anzuhängen. Auch wenn diese Tests aufgrund ihres Status als Zusatzuntersuchungen keineswegs hinreichend die Problematik zu erfassen vermögen, mit anderen Problemen der Datenqualität behaftet sein mögen (vgl. dazu Kapitel 3.2 und 4.) und die Hypothesen dem allgemeinen Design angepasst werden mussten, so erscheinen uns die Befunde dennoch berichterstattenswert und erfüllen hoffentlich die eine Funktion, weitere und bessere Tests anzuregen.

3.1 Die Hypothesen

Studien zur sozialen Erwünschtheit sowie Methodenuntersuchungen legen nahe, das telefonische Interview im Gegensatz zum persönlichen als eine Erhebungsmethode zu verstehen, die weniger Verzerrung provoziert. Mindestens auf aggregiertem Niveau lässt sich diese These anhand der Abstimmungs-Daten (VOX-Daten) testen, weil im Jahre 1995 die Umstellung auf dezentral durchgeführte, telefonische Interviews stattgefunden hat. Folglich lautet unsere Hypothese:

- 1.1. *Mit dem Übergang von persönlichen zu telefonischen Interviews wird der Durchschnittswert im Overreporting signifikant tiefer.*

Darüber hinaus haben wir die These der sozialen Erwünschtheit in einem Feldexperiment getestet. Wenn soziale Erwünschtheit das OR erklärt, dann sollte mit einem entsprechenden Wording der Frage das OR reduzierbar sein. Deshalb haben wir die Hypothese aufgestellt:

- 1.2. *Wenn im Einleitungstext zur Teilnahmefrage darauf hingewiesen wird, dass Nichtbeteiligung üblich und verbreitet ist, dann gibt es signifikant weniger Overreporting als bei einer Frage ohne diesen Hinweis.*

Abnehmende Parteibindungen und Partizipationsraten gelten gemeinhin als Indikator für Individualisierungsprozesse. Im Kontext der schweizerischen Referendumsdemokratie äußert sich dieser Prozess nicht in Form einer linearen Abnahme der Beteiligung, sondern in stark fluktuierenden Teilnahmequoten. Nur wenige Stimmbürger gehen aus staatsbür-

gerlichem Pflichtbewusstsein regelmäßig zur Urne, umso mehr dafür selektiv und je nach Betroffenheit. Daher formulieren wir die Hypothese:

- 1.3. Während die Stimmbeteiligung stark schwankt, nimmt das Overreporting im Verlaufe der Zeit ab.*

In Bezug auf die Erinnerungsleistung und das source monitoring haben wir uns vor allem an den Begründungen von Abelson et al. (1992) sowie Belli et al. (1994 bzw. 1999) orientiert. So lässt sich das auf der individuellen Ebene formulierte Argument, wonach memory confusing vor allem bei Personen auftreten sollte, welche öfters teilnehmen (Abelson et al. 1992: 142), auf der Makro-Ebene des schweizerischen Kontexts wie folgt formulieren: Da sehr viel mehr nationale Abstimmungen als nationale Wahlen durchgeführt werden und erstere auch noch in recht enger Kadenz sowie mit mehreren Vorlagen erfolgen, sollte das source confusing vor allem bei Sachabstimmungen ein Problem darstellen. Entsprechend haben wir die Hypothese aufgestellt:

- 2.1. Bei Abstimmungen ist das Overreporting deutlich höher als bei Wahlen.*

Für den experimentellen Test haben wir die Argumentation von Belli et al. (1999: 91) aufgenommen, wonach soziale Erwünschtheitsprobleme vor allem kombiniert mit Erinnerungsproblemen auftreten.

- 2.2. Wenn in einer Teilnahmefrage auch die Erinnerungsproblematik antizipiert wird, dann gibt es signifikant weniger Overreporting als bei einer Frage ohne diesen Hinweis.*

Zudem haben wir in Anlehnung an Wright (1993) sowie Belli et al. (1999) den Faktor Zeit in Betracht gezogen. Je weiter weg das zu berichtende Ereignis – so die allgemeine Annahme dahinter –, umso stärker kommen response sets zum Tragen.¹⁹ Daher haben wir die gleiche Formulierung, die in der zweiten Befragung verwendet wurde, in einem dritten Test nochmals verwendet. Abgesehen vom anderen Thema und einem neuen Sampling hat sich dabei vor allem der Zeithorizont geändert: Die Frage wurde nun nicht unmittelbar nach der Abstimmung, sondern rund zwei Monate danach gestellt. Die Hypothese dazu lautete:

- 2.3. Der Effekt der Frageformulierung ist bei der Befragung mit einem größeren Zeitabstand größer als bei der Befragung unmittelbar nach dem Urnengang.*

19 Es gilt aber zu beachten, dass das von Wright beobachtete response set ein Springen auf den Bandwagon beinhaltet, bei Belli et al. ist es das Overreporting.

Schließlich haben wir uns – soweit es die Daten zuließen – auch den Profilen gewidmet. Dabei sind wir aufgrund der sich abzeichnenden Tendenz (vgl. Abschnitt 2.2) und der theoretisch recht konsistenten Begründungen für das Experiment 1 von folgender Annahme ausgegangen:

- 3.1. *Auf das Wording zur Abfederung der sozialen Erwünschtheit reagieren vor allem die höher Gebildeten und politisch Interessierten/Involvierten mit weniger Overreporting.*

Im Gegensatz dazu sollte bei einem Wording, das den Vorschlägen von Belli et al. folgt, kein eindeutiges Profil erkennbar sein. Denn ein Wording, welches Erinnerungsprobleme und soziale Erwünschtheit kombiniert, kann zum Beispiel abnehmendes OR bei den weniger Gebildeten (Erinnerung) und höher Gebildeten (Erwünschtheit) gleichzeitig produzieren. Und ähnliches gilt, wenn man Gedächtnisprobleme als solche in Betracht zieht: Aufgrund der Theorie ist zu erwarten, dass reine Erinnerungsprobleme bei formal weniger Gebildeten und weniger Involvierten auftreten, während das source confusing jedoch bei formal höher Gebildeten und politisch Involvierten erwartet wird. Entsprechend vermuten wir:

- 3.2. *Auf ein Wording nach der Vorlage von Belli et al. reagieren Befragte an beiden Polen – höher Gebildete und stärker Involvierte ebenso wie weniger Gebildete und politisch kaum Interessierte – mit weniger Overreporting.*

3.2 Die Daten

Unsere Analysen basieren zum einen auf den „ex-post“ Analysen zu den eidgenössischen Urnengängen (VOX-Analysen) und Wahlen (Swiss Election Study) sowie auf drei „split-half“ Feldexperimenten.

VOX-Befragungen: Die Daten basieren auf einer zweistufigen (Haushalt – Personen) und geschichteten (Sprachregionen – Haushalte) ‚Zufallsstichprobe‘²⁰ von jeweils rund 1.000 Personen (ca. 700 Deutschschweiz, 240 Westschweiz, 60 italienische Schweiz). Die Befragungen wurden bis 1995 persönlich, ab Juni 1995 telefonisch, jedoch dezentral, durchgeführt. Die VOX sind vom Bund unterstützt, die Feldarbeit wird vom Meinungsforschungsinstitut GfS in Bern realisiert; die Universitäten Bern, Genf und Zürich zeich-

20 Die Auswahl der Personen erfolgt über die Geburtstagsmethode. Allerdings wird bei der GfS nicht alles dem Zufall überlassen, die Verantwortlichen der Stichprobe geben an: „Da sich auf der individuellen Ebene aber unterschiedlich starke Probleme mit der Erreichbarkeit ergeben, fügen wir zur Verbesserung der Stichprobenqualität Maximalgrößen bei“ (Technischer Bericht VOX Nr. 73: 13).

nen rotierend für Fragebogen und Analyse verantwortlich. Ab VOX Nr. 63 (Juni 1998) liegen Methodenberichte vor; die durchschnittlich berechnete Ausschöpfung beträgt 20.3 Prozent (Basis: bereinigte Bruttostichprobe, ohne stichprobenneutrale Ausfälle). Wir danken Lukas Golder von der GfS für die Kooperation.

Selects: Bis zu den Wahlen 1991 wurden die schweizerischen Wahlanalysen im Rahmen der VOX-Befragungen realisiert und folgten weitgehend diesen Stichproben. Seit 1995 werden die Nachwahlbefragungen von Selects (Swiss Electoral Studies), einem Zusammenschluss der Universitäten Bern, Genf und Zürich, organisiert. Basierend auf einer zweistufigen Zufallsstichprobe (Haushalt – Person), werden jeweils in den drei Wochen nach dem Wahlwochenende die telefonischen Befragungen durchgeführt (in der ganzen Schweiz, mit kantonalen Aufstockungen). 1995 wurden von der IPSO Sozial-, Marketing- und Personalforschung und vom Forschungsinstitut IHA-GfM zusammen über 7.000 Interviews realisiert. 1999 und 2003 war das Markt- und Sozialforschungsinstitut Link für die Befragung zuständig (3.258 resp. 5.891 Interviews). Die Ausschöpfung betrug 40.5, 51.3 und 46.9 Prozent (1995-2003).

Befragung 1: Das erste Feldexperiment wurde von DemoSCOPE zwischen dem 30. Januar und dem 5. Februar 2004 im Rahmen einer SNF-Studie über Informationsprozesse durchgeführt – also im Vorfeld der Abstimmungen über den Gegenvorschlag zur Avanti-Initiative, über die Verwahrungs-Initiative und über die Änderung des Mietrechts vom 8. Februar 2004. Die kurze Feldlaufzeit und das Interesse an Informationsprozessen von Teilnehmewilligen prägte die Stichprobe: Befragt wurden 1.004 Stimmberechtigte aus der Deutschschweiz; die Ausschöpfung der random/random-Stichprobe betrug 35.3 Prozent. Die Befragung fand vom zentralen Telefonlabor aus statt.

Befragung 2: Das zweite Feldexperiment wurde wiederum von DemoSCOPE eine Woche nach den Abstimmungen vom 16. Mai 2004 durchgeführt, bei denen es um folgende Vorlagen ging: 11. AHV-Revision, Erhöhung der Mehrwertsteuer zugunsten von AHV/IV und Steuerpaket. Hier wurden 700 Personen in der Deutschschweiz in einem random/quota-Verfahren befragt. Das Total der verwendeten Ausgangsadressen betrug 3.472; 837 der Ausfälle waren neutral, 1.935 könnten verzerrend wirken (1.095 nicht erreichbar; 839 Verweigerungen); womit sich eine ‚Ausschöpfung‘ von 26.6 Prozent ergab.

Wir danken Frau Tschurennev von DemoSCOPE für die Begleitung.

Befragung 3: Das dritte Feldexperiment behandelte wiederum die Abstimmung vom 8. Februar 2004, wurde jedoch 77 Tage nach der Abstimmung realisiert. Es handelte sich um eine Methodenstudie (Unit-Nonresponse) im Auftrag des SIDOS und in Kooperation mit dem Bundesamt für Statistik, welches für das Sampling zuständig war. Bei Subsample 1

wurde nach der ersten Verweigerung ein Incentive verschickt und es wurden Konvertierungsinterviews durchgeführt. Für Subsample 2 fanden Konvertierungsinterviews ohne Incentive statt; Subsample 3 wurde gemäß üblichem Standard, wo sofort die Zielperson ausgewählt wird, durchgeführt. Aufgrund der Testanlage wurden zwei Kontakte hergestellt: zwischen dem 17.3.-3.4. (Haushalt) und vom 26.4.-15.5.2004 für die eigentlichen Personeninterviews: In diese Phase fiel somit auch das Experiment. Vom ursprünglichen Adressenset (750) konnten 369 Interviews realisiert werden; womit sich nach Abzug der stichprobenneutralen Ausfälle auf der Ebene von Haushalt und Personen ($n = 86$) eine Ausschöpfung von 55.6 Prozent ergibt. Wir danken dem Studienleiter, Dominique Joye, für den großzügigen Zugang zu dieser Erhebung.

Die Daten stammen aus Erhebungen, die – mit einer Ausnahme – für andere Zwecke realisiert wurden. Daher müssen wir insbesondere auch bei den unabhängigen Variablen auf unterschiedliche Operationalisierungen zurückgreifen. Das zweite und dritte Feldexperiment wurde fast zeitgleich realisiert; die Nummerierung folgt der Argumentationslogik in diesem Artikel. Über die genaue Frageformulierung gibt Tabelle 1 Auskunft. Die Zuweisung zu den Experimentalgruppen war in allen Fällen randomisiert. Der Einfachheit halber sprechen wir von Experimentalgruppe dort, wo ein gezieltes Wording vorgenommen wurde und von Kontrollgruppe dort, wo eine Standardfrage verwendet wurde.

Tabelle 1 Frage- und Antwortformulierungen der Feldexperimente

1. Befragung Experimentalgruppe Bei solchen Abstimmungen geht normalerweise mehr als die Hälfte der Stimmberechtigten nicht an die Urne, es gibt schließlich noch anderes als Politik. Wie ist das bei Ihnen, nehmen Sie voraussichtlich an der Abstimmung vom 8.2. teil oder nicht, oder haben Sie sich noch nicht entschieden? ja, habe schon abgestimmt ja, will abstimmen gehen nein, will nicht abstimmen gehen ich habe noch nicht entschieden weiß nicht keine Angabe		Kontrollgruppe (Standardfrage) Wie ist das bei Ihnen, nehmen Sie voraussichtlich an der Abstimmung vom 8.2. teil oder nicht, oder haben Sie sich noch nicht entschieden? ja, habe schon abgestimmt ja, will abstimmen gehen nein, will nicht abstimmen gehen ich habe noch nicht entschieden weiß nicht keine Angabe	
2. Befragung Experimentalgruppe	3. Befragung Experimentalgruppe	2. Befragung Kontrollgruppe (Standardfrage)	3. Befragung Kontrollgruppe (Standardfrage)
Am Wochenende fand eine Abstimmung über das Steuerpaket, die 11. AHV-Revision und die Erhöhung der Mehrwertsteuer zu Gunsten der AHV/IV statt.	Vor etwa knapp zwei Monaten, am 8. Februar fand eine eidgenössische Volksabstimmung über die Verwahrungsinitiative, den Avanti-Gegenvorschlag und eine Änderung des Mietrechts statt.	Am Wochenende fand eine Abstimmung über das Steuerpaket, die 11. AHV-Revision und die Erhöhung der Mehrwertsteuer zu Gunsten der AHV/IV statt.	Vor etwa knapp zwei Monaten, am 8. Februar, fand eine eidgenössische Volksabstimmung über die Verwahrungsinitiative, den Avanti-Gegenvorschlag und eine Änderung des Mietrechts statt.
Versuchen Sie sich nochmals kurz zu erinnern. Haben Sie an der Abstimmung vom 16. Mai teilgenommen, wollten sie ursprünglich teilnehmen, haben sie nicht teilgenommen oder können Sie sich nicht mehr erinnern?	Versuchen Sie sich nochmals kurz an diese Zeit zu erinnern. Haben Sie an der Abstimmung vom 8. Februar teilgenommen, wollten sie ursprünglich teilnehmen, haben sie nicht teilgenommen oder können Sie sich nicht mehr erinnern?	Wie war das bei Ihnen, haben Sie an der Abstimmung vom 16. Mai teilgenommen oder nicht?	Wie war das bei Ihnen, haben Sie an der Abstimmung vom 8. Februar teilgenommen oder nicht?
Ja, habe teilgenommen	Ja, habe teilgenommen	ja nein	ja nein
Wollte ursprünglich teilnehmen, war dann aber verhindert	Wollte ursprünglich teilnehmen, war dann aber verhindert	(Nicht vorgelesen):	(Nicht vorgelesen):
Wollte ursprünglich teilnehmen, aber konnte mich nicht entscheiden	Wollte ursprünglich teilnehmen, aber konnte mich nicht entscheiden	keine Angabe	keine Angabe
Nein, habe nicht teilgenommen	Nein, habe nicht teilgenommen		
(Nicht vorgelesen): Weiß nicht mehr	(Nicht vorgelesen): Weiß nicht mehr		
Nicht stimmberechtigt			

3.3 Befunde

Tabelle 2 Datensätze und Overreporting im Überblick – Angaben in Prozent

Datensatz	n	Selbst-deklarierte Teilnahme/ Absicht ³⁾	Offizielle Teilnahme	Overreporting (d)
66 Volksabstimmungen (VOX-Befragungen 1977-2004) ¹⁾	ca. 1.003	58.1	42.6	15.5
7 Wahlen (VOX-Befragungen und Selects 1995 ff.) ²⁾	ca. 2.850	63.5	47.6	15.9
1. Befragung vor der Abstimmung vom 8.2.2004	1.000	86.5 ⁴⁾	45.0	41.5
2. Befragung nach der Abstimmung vom 16.5.2004	700	75.9	50.2	25.7
3. Befragung 2 Monate nach der Abstimmung vom 8.2.2004	369	67.1	45.0	22.1

¹⁾ Es handelt sich um Durchschnittswerte basierend auf den substanziellen und gerundeten Zahlen für jede VOX

²⁾ Durchschnittswerte der Nachwahlbefragungen 1971-2003, ohne 1983

³⁾ in den nachfolgenden Tabellen wird der Einfachheit halber von Teilnahme gesprochen

⁴⁾ inkl. „ja, ich habe die Absicht teilzunehmen“

Als erstes zeigt sich in Tabelle 2, dass das Phänomen des OR in allen verwendeten Datensätzen zu beobachten ist, die Vorwahlbefragung aber ähnlich wie schon von Jackman (1999: 18) aufgezeigt, besonders anfällig dafür ist. Bei den „ex-post“ Analysen zu den eidgenössischen Urnengängen ist das OR zwar vergleichsweise gering, bewegt sich aber ziemlich genau auf dem Niveau, das in den Wahlstudien anderer Länder zu beobachten ist.

Wenn wir gemäß Hypothese 1.1. den Blick etwas stärker auf diese Abstimmungserhebungen fokussieren, dann zeigt sich zudem: Hinter dem Mittelwert 15.5 versteckt sich eine hohe Spannweite von 17 Prozentpunkten. Das OR kann bis auf acht Prozentpunkte sinken und bis auf 25 Prozentpunkte hinaufklettern. Ein abnehmender Trend ist jedoch nicht zu beobachten und Tabelle 3 illustriert, dass der Übergang von persönlichen (face-to-face) zu telefonischen Interviews keine signifikante Reduktion mit sich gebracht hat. Verknüpft mit der Tatsache, dass wir es nie mit einem Underreporting zu tun haben, kann man aus diesen Daten noch keinen Beweis gegen die soziale Erwünschtheitstheese ablesen. Vielleicht haben wir auch einfach einen Hinweis dafür, dass es gerade im Rahmen von schnell durchgeführten Telefonbefragungen schwierig ist, dem Phänomen mit einer Maßnahme auf der Ebene des Wording zu begegnen.

Tabelle 3 Overreporting und Erhebungsmethode (Interviewstil) in den VOX-Analysen

	Mittelwert Face to face	Mittelwert telefonisch	Quadrat- summe	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	15.4	15.4	0.005	1	0.005	0.001	0.982

Mit Hilfe des „split-half“ Experiments in der 1. Befragung konnten wir die These mit einem Zugang von höherer interner Validität testen. Ein Einleitungstext zur Abfederung der sozialen Erwünschtheit sollte das OR in der Experimentalgruppe signifikant verringern. Das in Tabelle 4 präsentierte Datenmaterial macht allerdings deutlich, dass die Nullhypothese zur Hypothese 1.2. nicht widerlegt werden kann. Damit bestätigen sich die Befunde aus den USA. Eine leicht veränderte andere Formulierung mit Antizipation der sozialen Erwünschtheit vermochte weder auf aggregiertem Niveau der ANES noch in experimentellen Wording-Tests einen Effekt zu produzieren, wobei wir bei der vorliegenden Untersuchungsanlage über die Gründe nur spekulieren können. Probleme der Vorwahlbefragung und des Sampling könnten kombiniert mit der Tatsache, dass der experimentelle Stimulus im Rahmen eines telefonischen Interviews schlicht zu schwach ist, dafür verantwortlich sein.

Tabelle 4 Effekt des Wording (soziale Erwünschtheit) auf das Antwortverhalten

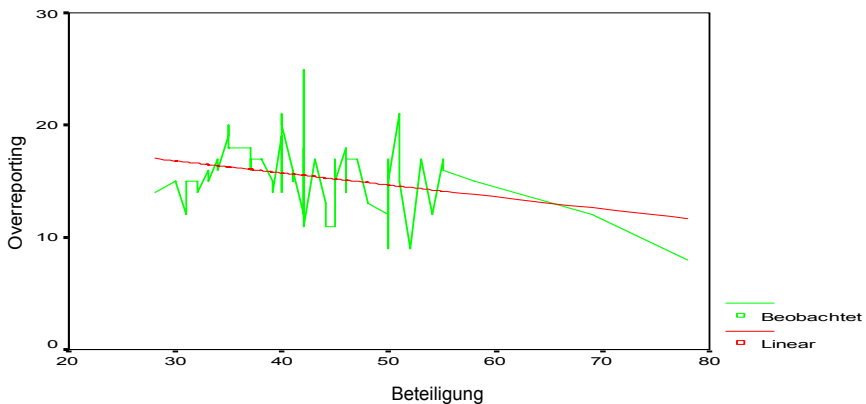
Test	Treatment	Teilnahme ja ¹⁾ in Prozent	Teilnahme nein in Prozent	n	Signifikanz ²⁾
1. Befragung vor der Abstimmung Februar	EG	91.0 (436)	9.0 (43)	479	0.475
	KG	90.7 (429)	9.3 (44)	473	

Lesebeispiel: Bei der ersten Befragung haben in der Experimentalgruppe (EG) mit dem SD-Wording (vgl. Tabelle 1) 91.0 Prozent angegeben teilzunehmen, in der Kontrollgruppe mit der Standardfrage waren es 90.7 Prozent. Der Unterschied zwischen den Gruppen ist somit nicht signifikant.

¹⁾ inkl. Teilnahmeabsicht.

²⁾ Exakter Fisher-Test, einseitig.

Auch die Hypothese 1.3. lässt sich nicht bestätigen. Die Beteiligung an Abstimmungen ist zwar in der Tat stark schwankend und die Kurve dokumentiert den selektiven Urnengang in der Schweiz. Das Overreporting nimmt aber im Verlaufe der Zeit keineswegs ab ($Rsq = 0.002$; $df = 64$; $F = 0.13$; $Sig. = 0.724$; $b1 = -0.000$). Allerdings impliziert die ausbleibende Abnahme auch nicht, dass die konträre These von Burden (2003 und 2000) zutrifft, wonach das OR im Verlaufe der Zeit sogar zunimmt. Hingegen stellen wir einen schwachen, aber eindeutig negativen Zusammenhang zwischen der Stimmbeteiligung und dem OR fest ($Rsq = 0.102$; $df = 64$; $F = 7.27$; $Sig. = 0.009$; $b1 = -0.107$).

Grafik 1 Overreporting nach Stimmbeteiligung

Insofern können wir Granberg & Holmberg (1991: 457) bestätigen, welche darauf hinweisen, dass die Validität der Partizipationsfrage eine Funktion der hohen Teilnahmerate darstellt. Das wiederum deutet – im Sinne eines plausiblen, aber keineswegs kausalen Arguments – darauf hin, dass wir es bei der VOX-Analyse weniger mit einem Mess- sondern vielmehr mit einem Sampling-Problem zu tun haben. Im Moment, wo die „population at risk“ – die Nichtwähler – in der Population weniger stark vertreten ist, scheint auch deren Unterrepräsentanz weniger problematisch.

Aus theoretischer Sicht ist diese Beobachtung eines negativen Zusammenhangs allerdings schwierig zu erklären. Im Fall eines positiven Zusammenhangs könnte man mit Rückgriff auf die soziale Erwünschtheit behaupten, dass bei einer höheren Stimmbeteiligung das OR steigt, weil Befragte – sofern sie das Wissen über die Teilnahme haben – mit einer Art Konsensus-Heuristik nach dem Motto „wenn so viele gingen, dann muss das wohl ein wichtiger Urnengang gewesen sein“ reagieren. In etwa so argumentieren auch Karp & Brockington (2005) und stützen ihre These mit gepoolten Individualdaten aus Validitätsstudien von fünf Ländern. Für das hier vorliegende Muster mit sinkendem OR lässt sich indessen mit der Erwünschtheitstheorie kaum eine Erklärung konstruieren. Somit werden bei diesem Analyseschritt nicht nur die Grenzen eines Zugangs mit Aggregatdaten offenkundig, sondern es wird deutlich, dass Mess- und Sampling-Probleme gleichzeitig untersucht werden sollten.

Ausgehend von den Ausführungen zum source monitoring haben wir die Hypothese aufgestellt, bei den schweizerischen Abstimmungsanalysen müsse das OR ausgeprägter sein als bei den Wahlen. Ein Vergleich der entsprechenden Mittelwerte vermag diese Annahme

nicht zu bestätigen (vgl. Tabelle 5). Werden sämtliche vorliegenden Studien miteinander verglichen, unterscheiden sich die Mittelwerte nicht (15.4 Abstimmungen; 15.8 Wahlen). Bei einer Reduktion auf jene Fälle, die mit vergleichbaren Sampling-Prozeduren erstellt wurden (ab 1995) stellen wir das Gegenteil der Hypothese fest. Das OR ist bei den Wahlen ausgeprägter. Aufgrund dieses einfachen bivariaten Zugangs und der Tatsache, dass wir das source monitoring nicht direkt erfassen können, ist das noch kein starker Gegenbeweis zur These. Aber er lädt zumindest dazu ein, sich über die Bedingungen des source confusing und möglicher Gedächtniseffekte Gedanken zu machen. Unsere Annahme war, dass Abstimmungen aufgrund der hohen Frequenz anfällig für OR sind. Die vorliegenden Daten legen indessen nahe, davon auszugehen, dass – wenn dieser kognitive Faktor überhaupt eine Rolle spielt – nicht Häufigkeit mit Konfusion einhergeht und Konformität produziert, sondern eher Seltenheit oder geringer Stellenwert mit Vergesslichkeit einhergeht und Konformität zur Folge hat. Damit hat vielleicht Presser (1984: 353) mit seiner Vermutung Recht behalten: „But the process may also have been a more subtle one, affecting respondents’ memories. This seems especially likely for an action like voting [gemeint ist das Wählen], which is neither central to the lives of most people nor performed frequently.”²¹

Tabelle 5 Overreporting nach Typ des Urnengangs (alle und ab 1995)

	Mittelwert Abstimmungen	Mittelwert Wahlen	Quadrat- summe	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen (alle)	15.4	15.8	1.261	1	1.261	0.082	0.775
Zwischen den Gruppen (ab 1995)	15.4	20.2	64.24	1	64.24	6.259	0.018

Ohnehin stößt man aber bei einem kognitionspsychologischen Konstrukt mit Aggregatdaten an analytische Grenzen, weshalb wir uns lieber den Individualdaten widmen wollen. Für den Test der Hypothesen 2.2. und 2.3. haben wir in der 2. und 3. Befragung zwei Feldexperimente geschaltet, die sich – abgesehen vom sicher nicht zu vernachlässigenden thematischen Wechsel und dem unterschiedlichen Sampling (inkl. Ausschöpfungsquote) – im Zeitfaktor unterscheiden: Das zweite Experiment wurde unmittelbar nach der Mai-Abstimmung realisiert; das dritte mehr als zwei Monate nach der Februar-Abstimmung.

21 In der Tat teilen die befragten Stimmberechtigten in den VOX-Analysen immer wieder mit, Abstimmungen seien wichtiger als Wahlen.

Bei der Befragung nach den Abstimmungen im Mai hat das unterschiedliche Wording keinen Effekt; die beiden Splits unterscheiden sich nicht signifikant (vgl. Tabelle 6). Bei der Befragung nach den Abstimmungen im Februar hingegen, die erst zwei Monate später stattgefunden hat, ist der erwartete Effekt zu beobachten: Mit der Experimentalfrage, welche die Erinnerungsproblematik (und die soziale Erwünschtheit) berücksichtigt, geben weniger Befragte an, dass sie an den Abstimmungen teilgenommen haben ($p = 0.088$). Ein Wording-Effekt in einer kurzfristig angelegten Befragung ist somit nicht effektiv; bei Hypothese 2.2. kann die Nullhypothese nicht verworfen werden. Hingegen haben wir – auch wenn der direkte Vergleich zwischen den zwei Befragungen aufgrund anderer Rahmenbedingungen nur beschränkt möglich ist – ein starkes Indiz für Hypothese 2.3. Liegt das Ereignis weiter zurück, so kann mit einem entsprechenden Wording ein positiver Effekt erzielt werden; das Overreporting nimmt ab.

Tabelle 6 Effekt des Wording (source monitoring) auf das Antwortverhalten

Test	Treatment	Teilnahme ja in Prozent	Teilnahme nein ¹⁾ in Prozent	n	Signifikanz ²⁾
2. Befragung nach der Abstimmung vom 16. Mai	EG	77.4 (270)	22.6 (79)	349	0.219
	KG	74.6 (261)	25.4 (89)	350	
3. Befragung 2 Monate nach der Abstimmung vom 8. Februar	EG	63.9 (108)	36.1 (61)	169	0.088
	KG	71.7 (104)	28.3 (41)	145	

¹⁾ inkl. „wollte ursprünglich teilnehmen, war dann aber verhindert“ bzw. „konnte mich aber nicht entscheiden“

²⁾ Exakter Fisher-Test, einseitig

Die Wirkung scheint jedoch nicht überragend.²² Das Overreporting bleibt selbst in der Experimentalgruppe bestehen (63.9 Prozent selbst-deklarierte Teilnehmende gegenüber 45 Prozent in der offiziellen Statistik). Zudem ist die provozierte Prozentpunkt-Differenz zwischen der Experimental- und der Kontrollgruppe mit 7.8 nicht sehr ausgeprägt. Denn im Vergleich dazu wird in unserer Referenzstudie bei einem Design mit kleinerem time lag (56 Tage) eine starke Differenz von 17.6 Prozentpunkten provoziert (Belli et al. 1999: 102). Während dort in der Kontrollgruppe mit der ANES-Standardfrage immer noch 74.9 Prozent angaben, sie hätten an den Wahlen teilgenommen, sinkt der Anteil in der Experimentalgruppe auf 57.3 Prozent. Eine Erklärung für diesen starken Effekt liegt vermutlich im Wording des Experiments:

22 Der Begriff „scheint“ ist hier durchaus am Platz, weil wir Effekte des Sampling, das vielleicht eine höhere Beteiligungsrate erklärt, nicht kontrollieren können und keine Validierung der Daten möglich ist.

Tabelle 7 Wording im Experiment von Belli et al. (1999: 106)**Experimental Wording, National Study**

The next question is about the elections in November. In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time.²³ We also sometimes find that people who thought that they had voted actually did not vote. Also, people who usually vote may have trouble saying for sure whether they voted in a particular election. In a moment I am going to ask you whether you voted on Tuesday, November 5th, which was {time fill} ago. Before you answer, think of a number of different things that will likely come to mind if you actually did vote this last past election day; things like whether you walked, drove, or were driven by another person to your polling place {pause}, what the weather was like on the way {pause}, the time of day that was {pause}, and people you went with, saw, or met while there {pause}. After thinking about it, you may realize that you did not vote in this particular election {pause}. Now that you've thought about it, which of these statements best describes you?"

I did not vote in the November 5th election (1);

I am sure I voted in the November 5th election (4)

I thought about voting this time but didn't (2);

I voted by absentee ballot (volunteered, 7).

I usually vote but didn't this time (3)

Während diese Formulierung darauf abzielt, soziale Erwünschtheit zu reduzieren (kursiv) und vor allem das Retrievement zu erleichtern (unterstrichen), indem mit ‚kognitiven Bildern‘ die relevanten Informationen im Gedächtnis verfügbar gemacht werden, könnte diese Formulierung allerdings auch eine Art Präzisions- oder Kontrolleffekt auslösen (Formulierungen ohne Hervorhebung). Wenn wir nämlich in Betracht ziehen, dass die Befragten von Vorgaben in einem Fragebogen ‚lernen‘ (Schwarz & Scheuring 1992), dann sind mindestens zwei weitere Erklärungen für die Effektivität des von Belli et al. gewählten, aber, zumindest in telefonischen Interviews ungeeigneten Wordings möglich: Zum einen könnte es sein, dass mit der ausführlichen Frage den Befragten eindringlich genug kommuniziert wird, dass hier nun große Sorgfalt verlangt ist. Dann hätten wir es vielleicht mit jenem Präzisionseffekt zu tun, auf den bereits eine Koautorin von Abelson aufmerksam gemacht hat (Loftus et al. 1990). Zum anderen ist nicht auszuschließen, dass die Befragten nach dieser langen Einleitung vermuten, die Frage, wie man zum Wahllokal gekommen sei, werde später im Interview nochmals aufgenommen. Oder die Konkretheit der Situationsschilderung vermittelt den Interviewten sogar der Eindruck, man hätte Zugang zu Kontrolldaten (wenn auch nur Freunde, die die Wählenden vermeintlich begleitet haben). Vielleicht haben also Belli und sein Team ein Treatment entwickelt, das – unbeabsichtigt aber effektiv – einen Gedanken aufnimmt, den Abelson et al. (1992: 151) formuliert haben und den klassischen „bogus pipeline studies“ entnommen ist (Jones & Sigall 1971). Demnach können Antwortverzerrungen durch soziale Erwünschtheit vor

23 Der Text bis hierhin entspricht der ANES-Standardfrage.

allem dann reduziert werden, wenn die Befragten meinen, unaufrichtige Antworten könnten durch Kontrollen entdeckt werden (Tourangeau et al. 2000: 277).²⁴

Abschließend möchten wir die zentralen Befunde zu den Profilen jener zusammenfassen, die auf ein Treatment zur sozialen Erwünschtheit und/oder zu Gedächtniseffekten reagieren (vgl. Tabelle 8).

Tabelle 8 Subgruppenanalyse und Merkmale mit Effekt

Merkmal/Subgruppe	Wo/ Nummer der Befragung	Effekte: Differenz in Prozentpunkten	
		Wie/ Abnahme Ja Anteil	Wie/ Zunahme Ja Anteil
Mittleres Alter (35- bis 54-Jährige)	3	14.2* (117)	
<i>Bildung</i>			
Primar- und Sekundarschule	2	14.6 * (110)	
Sekundar-, Real-, Bezirksschule	2	17.5 * (87)	
Berufsschule	3 / 2	19.1*** (177)	7.3* (337)
Alles außer Primar- und Sekundarschule	3	12.8** (254)	
Alles außer Uni/ETH	3	8.1* (287)	
Berufsschule bis Uni	2		5.8* (583)
Teilzeitarbeit	3	17.2* (64)	
Vertrauen in PolitikerInnen			
Wenig (sehr wenig und wenig)	3	13.9* (128)	
Sehr wenig	3	23.9* (45)	
<i>Interesse für Politik</i>			
kaum	3	23.1** (74)	
Gar nicht	1		43.2 ** (35)
<i>Wahrnehmung Politik als kompliziert</i>			
ziemlich häufig	3	29.2** (57)	
häufig (ziemlich häufig und häufig)	3	23.6 *** (106)	
Effektivität: wenig Politiker kümmern sich	3	12.7* (120)	
Keine Petition unterzeichnet	3	14.9** (208)	
Nie Diskussionen über Politik	1		26.7** (65)
Beitritt in Partei vorstellbar	2		22.7* (110)
Debatten über Abstimmung verfolgt	2		9.2 * (226)
Stichprobe 1: mit Incentives	3	22.2 (96) **	

Lesebeispiel: Bei der Variable ‚Alter‘ konnte einzig unter den 35- bis 54-Jährigen eine signifikante Reaktion auf das Wording registriert werden – und zwar in Befragung 3: Während in der ‚Kontrollgruppe‘ 73.6 Prozent angeben, sie hätten teilgenommen, sind es in der Experimentalgruppe 59.4 Prozent. Somit ergibt sich die in der Tabelle ausgewiesene Differenz von 14.2 Prozentpunkten. In die entsprechende Vierfeldertabelle sind insgesamt 117 Fälle eingeflossen.

Signifikanz: Exakter Fisher-Test, 1-seitig: *** $p < .01$, ** $p < .05$, * $p < .1$

24 Entsprechend metaphorisch ist der Begriff „bogus pipeline“. Die Experimentierer vermitteln den Befragten den Eindruck, sie hätten Zugang (eine Pipeline) zu den ‚wahren‘ Reaktionen bzw. Verhaltensweisen.

In Bezug auf die erste Befragung und die soziale Erwünschtheitstheorie bestätigt sich dabei, dass der schwache Nettoeffekt keineswegs durch Bewegungen in unterschiedliche Richtungen in den Subgruppen zustande kommt, sondern in der Tat schwach ist. Darüber hinaus, und das interessiert hier mehr, zeigt sich, dass die Empirie in doppelter Hinsicht gegen Hypothese 3.1. spricht: Erstens stellen wir nicht bei jenen Personen, die den aufrichtigen Wählern gleichen, eine Reaktion fest, sondern bei jenen Befragten, die sich offen als „gar nicht interessiert“ bezeichnen und nie über Politik diskutieren. Zweitens löst das Treatment das Gegenteil dessen aus, was beabsichtigt ist. Bei den Stimmberechtigten, die mit Politik gar nichts anfangen können, nimmt das OR bei Verwendung der Erwünschtheitsfrage zu.²⁵ Wir sind hier wohl mit einem Mechanismus konfrontiert, der im Zusammenhang mit den „two-time frame“ Experimenten auch schon in Betracht gezogen wurde (Abelson et al. 1992: 142): Erst über das Wording scheint man hier auf die soziale Erwünschtheit der Stimmbeteiligung aufmerksam gemacht. Oder der explizite Hinweis auf verschiedene Gruppen von Wählern löst ein Distinktionsbedürfnis aus, ähnlich wie es auch dem Third-Person-Effekt inhärent ist. Damit löst die Frageformulierung nicht ein Problem, sondern schafft eines.

Und als wäre das der Komplexität nicht genug, stellen wir bei Befragung 2 fest, dass wir hier ähnliche negative Effekte des Wording haben, diesmal jedoch nicht die schlechter Integrierten, sondern Personen mit einem höheren Bildungsniveau (ab Berufsschule aufwärts) und politisch Involvierte (Beitritt in Partei vorstellbar, Debatten über Abstimmung verfolgt) noch mehr overreporten. Hypothese 3.2. ist somit dadurch widerlegt, dass die Bewegungen nicht wie vermutet in Richtung Abnahme des OR weisen. Hinzu kommt, dass wir bei Befragung 3 in der Tendenz und entgegen der Annahme ein Profil erkennen. Bei den soziodemographischen Variablen sind es eher die mittleren Kategorien, die reagieren. Bei den politischen Variablen sind es vor allem die Nichtinteressierten, die auf das Wording reagieren: Jene, die noch nie eine Petition unterzeichnet haben; jene, die glauben, sehr wenig Politiker würden sich darum kümmern, was die Wählenden denken; und Befragte, die nur mit zusätzlichen Incentives zur Teilnahme an der Befragung bewogen werden konnten. Dabei könnte die Reaktion der mittleren Kategorien bei den soziodemographischen Variablen durchaus in Einklang mit der source confusing stehen und impliziert vielleicht, die Annahme umzuformulieren: Gerade jene, die weder besonders stark noch besonders schwach ausgestattet sind, unterliegen einem source confusing. Insgesamt können wir aber vor allem festhalten: Der konsistent uniforme Effekt und die

25 Während unter der einfachen Teilnahmefrage ohne Einleitungssatz nur 23,5 Prozent der Befragten angeben, sie hätten teilgenommen, sind es unter der Bedingung des gezielten Wordings 66,7 Prozent, womit die in der Tabelle ausgewiesenen 43,2 Prozentpunkte Differenz zustande kommen. Der Chi-Quadratwert mit Yates Korrektur bestätigt die Signifikanz ($p = 0.026$).

vergleichsweise großen Prozentpunkt-Differenzen im dritten Feldexperiment unterstreichen die Annahme von Hypothese 2.3, dass, wenn das Ereignis weit zurückliegt, ein starker Wording-Effekt nachzuweisen ist. Welches die Ursachen dafür sind, bleibt aber solange unklar, bis Probleme der sozialen Erwünschtheit, der Erinnerung oder des source confusing nicht auch direkt erfasst werden.

4 Fazit

Das Overreporting politischer Partizipation – so zeigt die vorliegende Literaturdurchsicht ebenso wie das Länderbeispiel Schweiz – stellt ein offenkundiges Problem dar. Es wird in allen bisher untersuchten und etablierten Demokratien festgestellt und scheint nicht nur pankulturell, sondern auch auf hohem Niveau verbreitet zu sein. Auf aggregiertem Niveau werden leicht Differenzen von 15 Prozentpunkten erreicht und der Anteil in Validitätsstudien ‚entdeckter‘ unaufrichtiger Wähler kann bald 25 Prozent erreichen.

Dabei ist das Problem auch deshalb nicht zu unterschätzen, weil es mittlerweile zahlreiche Hinweise dafür gibt, dass das Misreporting in Befragungen zu einem „Overreporting der Sozialwissenschaftler“ führen kann. Wenn Overreporters überdurchschnittlich stark politisch interessiert sind, verleiten sie die Sozialwissenschaftler dazu, den Zusammenhang zwischen Interesse und Teilnahme in der Politik zu überschätzen. Unter Umständen bewegen wir uns also in einer Spirale des Common Sense, die uns nicht gefallen kann. Die einen orientieren ihre Antworten, die anderen ihre wissenschaftliche Modelle daran (Presser & Traugott 1992: 86). Zudem liefert die Analyse der VOX-Nachbefragungen einen weiteren Hinweis dafür, dass mit der Verbreitung der Telefoninterviews²⁶ entgegen der vielfach geäußerten Hoffnung nicht automatisch eine Abnahme des Overreporting einhergeht. Es scheint,²⁷ dass die vergleichsweise anonyme und unverbindliche Erhebungsform die starke Verankerung der gesellschaftlichen Norm „Abstimmen ist Bürgerpflicht“ nicht zu durchbrechen vermag. Umgekehrt kann es sehr gut sein, dass diese Erhebungsmethode Maßnahmen auf der Ebene des Wording erschwert. Die Tatsache jedenfalls, dass unser erstes Feldexperiment keine Wirkung und das Experiment von Belli und seinem Team mit einem sehr langen Einleitungstext dagegen viel Wirkung zeigte, kann auch damit zusammenhängen: Vielleicht verhindert die schnelle und spontane telefonische Erhebungsform, dass das Wording der Frageformulierung – in das die Forscher so viel investieren – in seiner Feinheit über-

26 Gemäß Statistik des Verbandes Schweizer Marketing- und Sozialforscher (SMS) wurden im Jahr 2000 66 Prozent der Befragungen telefonisch durchgeführt; auch in Deutschland dominiert diese Erhebungsmethode zu diesem Zeitpunkt (Gabler & Häder 2002).

27 Die zurückhaltende Formulierung ist angezeigt, weil Sampling-Effekte nicht ausgeschlossen werden können.

haupt wahrgenommen wird. Erst eine Frageformulierung, von der eigentlich jedes Lehrbuch aufgrund der Länge abrät, vermag vielleicht die notwendige Aufmerksamkeit der Befragten zu provozieren.

Insgesamt können also diese Maßnahmen nicht wirklich überzeugen. Zum ersten deshalb, weil sie zumeist Sampling und Unit-nonresponse-Probleme vernachlässigen, obwohl wir seit Brehm (1993) und den Studien von Voogt & Saris (2003) wissen, dass politische Partizipation und Befragungspartizipation korrelieren bzw. Sampling- und Messfehler in jeder Befragung in Kombination wirken können (Jackman 1999: 14). Hinweise für die Bedeutung der Sampling-Dimension konnten wir – oder mussten wir trotz Ermangelung entsprechender Daten – im Rahmen der oben stehenden Ausführungen mehrmals andeuten. Deshalb haben wir für die vielfach genutzten VOX-Analysen auch einen Strukturdatenvergleich vorgenommen. Die Daten verdeutlichen, dass die Stichproben in die erwartete Richtung verzerrt sind: potenzielle WählerInnen sind übervertreten (Angehörige der deutschen Schweiz, ältere Personen und höher Gebildete). Wir haben also auch hier starke Hinweise dafür, dass Overreporting auch auf Sampling-Aspekte zurückzuführen ist.

Unbefriedigend sind die besprochenen Maßnahmen zum zweiten auch deshalb, weil sie sich als riskantes Unterfangen entlarven können. Zuweilen scheinen sie nicht den Druck der Anpassung an die soziale Norm abzubauen, sondern lediglich diese soziale Norm (wieder) in Erinnerung zu rufen. Darüber hinaus sind die theoretischen Annahmen noch zu wenig klar und wir kennen die kausalen Prozesse zu schlecht, um experimentelle Manipulationen wirklich legitimieren zu können. Deshalb führt unseres Erachtens kein Weg daran vorbei, die möglichen Maßnahmen gegen ein Overreporting in zweierlei Richtungen weiterzudenken.²⁸

Erstens sollten die erklärenden Faktoren (soziale Erwünschtheit, Erinnerungsprobleme und source confusing) in Studien zum Overreporting parallel erfasst werden.²⁹ Denn nur mit einer solchen Parallelerhebung lässt sich klären, ob soziale Erwünschtheit (immer noch)³⁰ ein klares soziodemographisches Profil hat oder stärker vom Befragungskontext

28 Auf die mögliche Maßnahme des „random response“ wird hier nicht eingegangen, weil sie unseres Wissens vor allem in persönlichen Interviews eingesetzt wird, hier aber telefonische Befragungen im Vordergrund stehen. Für mögliche Testanlagen vgl. Locander et al. (1976: 270). Eine weitere Maßnahme würde darin bestehen, dass die kommerziellen Institute ihre Kontrollfragen und Prozeduren für die Erstellung von genauen Prognosen bekannt geben würden. Das kommt vielleicht einem frommen Wunsch gleich, und bis eine solche Studie realisiert ist, sei auf ein ähnliches Vorgehen kombiniert mit einer Validitätsstudie verwiesen: Miller (1952).

29 Zumindest in Bezug auf die soziale Erwünschtheit liegen ja mittlerweile auch deutschsprachige und ‚vergleichsweise‘ kurze Fragebatterien vor (Krebs 1991; Stöber 1999).

30 Zumindest ist das die Einschätzung von Krebs (1991: 7).

beeinflusst ist und damit als volatiler, aber messbarer Persönlichkeitszug zu verstehen ist. Zudem ließen sich Maße einer SD-Skala als Kontrollvariable in Modellschätzungen einsetzen, um zu demonstrieren, dass berechnete Zusammenhänge nicht ein Artefakt sozialer Erwünschtheit sind.

Zweitens möchten wir die These wagen, dass die Simulation einer Bogus-pipeline-Studie eine effiziente Maßnahme darstellen könnte. Denn nicht nur die Uniformität des Misreporting in Richtung Overreporting, sondern auch das Wording im Test von Belli et al. lässt vermuten, dass beim OR nicht nur Gedächtnisprobleme eine Rolle spielen, sondern vermutete Kontrolleffekte. Zudem wird häufig beobachtet, dass das Overreporting in selbst administrierten und damit schriftlichen Befragungen weniger hoch ist (Tourangeau et al. 295 ff.; Locander et al. 1976: 272). Vor diesem Hintergrund dürften selbst administrierte (schriftliche oder Online-) Befragungen, welche mit der Erhebungsmethode sowie der Frageformulierung gleichzeitig Verbindlichkeit und Kontrollmöglichkeiten suggerieren, wohl am ehesten helfen, das Overreporting zu reduzieren.³¹

Insgesamt erhält man damit den Eindruck, dass auch hier und bei uns kein Weg an mindestens einer größeren und aktuellen Validitätsstudie vorbeiführt, welche Misreporting und Nonresponse gleichzeitig analysiert. Denn nur ein solches Vorgehen erlaubt die Gleichsetzung von weniger Overreporting und mehr Akkuratheit. Zudem bilden Validitätsstudien die Grundlage für jene Maßnahmen, wie sie letzthin auch Jackman (1999) vorgeschlagen und Voogt (2000) angewendet haben: In bestehende Analysen werden Werte aus den Validitätsstudien zum OR als Proxies eingesetzt. Bisher wurde dies allerdings nur mit vereinzelt Validitätsstudien aus Holland,³² aus Australien und den USA getan. Wenn man aber den unterschiedlichen Zugang, die unterschiedlichen Analysemethoden sowie die divergenten Befunde derselben in Betracht zieht, dann ist dieses Vorgehen auch mit Risiken behaftet. Eine effektive Korrektur verlangt präzises Vorwissen, über das wir derzeit noch nicht verfügen. Daher scheint uns auch eine Meta-Analyse der zahlreich vorliegenden Studien ein wichtiger Schritt, und vielleicht inspiriert ja die vorliegende Literaturübersicht zur Durchführung derselben.

31 Dabei ist es wichtig darauf hinzuweisen, dass die Gegenmaßnahmen zum Overreporting sich von jenen gegen ein Underreporting unterscheiden sollten (Locander et al. (1976: 272).

32 Es handelt sich um die holländische Zaanstad-Studie anlässlich der nationalen Wahlen 1998 (Voogt & Saris 2003), die australische Validitätsstudie von 1993 (Vowles et al. 1995) sowie die Validitätsstudien von Traugott für die Jahre 1964-1988 (Traugott 1989).

Literatur

- Abelson, R. P., Loftus E. & Greenwald A. G. (1992). Attempts to Improve the Accuracy of Self-Reports of Voting. In J. M. Tanur (Ed.), *Questions About Questions. Inquiries into the Cognitive Bases of Surveys* (pp. 138-53). New York: Russell Sage Foundation.
- Abramson, P. R. & Claggett, W. (1986). Race-Related Differences in Self-Reported and Validated Turnout in 1984. *Journal of Politics* 48(2), 412-422.
- Adamany, D. & Shelley, M. C. (1980). Encore! The Forgetful Voter. *Public Opinion Quarterly*, 44(2), 234-240.
- Anderson, B. A. & Silver, B. D. (1986). Measurement and Mismeasurement of the Validity of Self-Reported Vote. *American Journal of Political Science*, 30(4), 771-785.
- Andersson, H. E. & Granberg, D. (1997). On the Validity and Reliability of Self-Reported Vote: Validity without Reliability? *Quality and Quantity*, 31, 127-140.
- Atkeson, L. R. (1999). 'Sure, I Voted for the Winner'. Overreport of the Primary Vote for the Party Nominee in the National Election Studies. *Political Behavior*, 21, 197-215.
- Belli, R. F., Traugott, M. W. & Beckmann, M. N. (2001). What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. *Journal of Official Statistics* 17(4), 479-498.
- Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63, 90-108.
- Belli, R. F., Traugott, S. & Rosenstone, S. J. (1994). *Reducing Over-Reporting of Voter Turnout: An Experiment Using a 'Source Monitoring' Framework*. Retrieved 2003, from <http://www.umich.edu/~nes/resources/other/techrpts/belli.htm>
- Bernstein, R. Chadha, A. & Montjoy, R. (2001). Overreporting [Us] Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65, 22-44.
- Brehm, J. (1993). *The Phantom Respondent. Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.
- Burden, B. C. (2003). Internal and External Effects on the Accuracy of NES Turnout: Reply. *Political Analysis*, 11(2), 193-195.
- Burden, B. C. (2000). Voter Turnout and the National Election Studies. *Political Analysis* 8(4), 389-398.
- Cahalan, D. (1968). Correlates of Respondent Accuracy in the Denver Validity Survey. *Public Opinion Quarterly* 32(4), 607-621.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Beverly Hills/London: Sage.
- Cassel, C. A. (2003). Overreporting and Electoral Participation Research. *American Politics Research*, 31, 81-92.

- Clausen, A. R. (1968). Response Validity: Vote Report. *Public Opinion Quarterly*, 32, 588-606.
- Crowne, D. & Marlowe, D. (1964). *The Approval Motive*. New York: Wiley.
- De Leeuw, E. D. & van der Zouwen, J. (1988). Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis. In P. B. Biemer (Ed.), *Telephone Survey Methodology* (pp. 283-99). New York: Wiley & Sons.
- Eilfort, M. (1994). *Die Nichtwähler. Wahlenthaltung als Form des Wahlverhaltens*. Paderborn: Schöningh.
- Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „Rational Choice“. In R. Wittenberg (Ed.), *Person – Situation – Institution – Kultur. Günter Büschges Zum 65. Geburtstag* (pp. 59-78). Berlin: Duncker & Humblot.
- Gabler, S. & Häder, S. (Eds.) (2002). *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland*. Münster: Waxmann.
- Granberg, D. & Holmberg, S. (1992). The Hawthorne Effect in Election Studies: The Impact of Survey Participation on Voting. *British Journal of Political Science*, 22, 240-248.
- Granberg, D. & Holmberg, S. (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35, 448-459.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Hill, K. Q. & Hurley, P. A. (1984). Nonvoters in Voters' Clothing: The Impact of Voting Behavior Misreporting on Voting Behavior Research. *Social Science Quarterly*, 65(1), 199-206.
- Hyman, H. H. (1945). Do They Tell the Truth? *Public Opinion Quarterly*, 8(4), 557-559.
- Jackman, S. (1999). Correcting Surveys for Non-Response and Measurement Error Using Auxiliary Information. *Electoral Studies*, 18, 7-27.
- Jones, E. E. & Sigall, H. (1971). The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude. *Psychological Bulletin*, 76, 349-364.
- Kaase, M. (1973). Die Bundestagswahl 1972: Probleme und Analysen. *Politische Vierteljahresschrift*, 14, 145-190.
- Kaase, M. & Bauer-Kaase, P. (1998). Zur Beteiligung an der Bundestagswahl 1994. In M. Kaase & H. D. Klingemann (Eds.), *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1994* (pp. 85-112). Opladen: Westdeutscher Verlag.
- Karp, J. A. & Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Over-Reporting Voter Turnout in Five Countries. *Journal of Politics*, 67(3), 825-840.
- Katosh, J. P. & Traugott, M. (1981). The Consequences of Validated and Self-Reported Voting Measures. *Public Opinion Quarterly*, 45, 519-535.
- Kleinhenz, T. (1995). *Die Nichtwähler. Ursachen der Sinkenden Wahlbeteiligung in Deutschland*. Opladen: Westdeutscher Verlag.
- Krebs, D. (1991). Was ist sozial erwünscht? Der Grad sozialer Erwünschtheit von Einstellungs-items. *ZUMA-Arbeitsbericht 91/18*.

- Lavies, R.-R. (1973). *Nichtwählen als Kategorie des Wahlverhaltens: Empirische Untersuchung zur Wahlenthaltung in historischer, politischer und statistischer Sicht*. Düsseldorf: Droste Verlag.
- Locander, W., Sudman, S. & Bradburn, N. B. (1976). An Investigation of Interview Method, Threat and Response Distortion. *Journal of American Statistical Association*, 71, 269-275.
- Loftus, E. F. et al. (1990). A Tale of Two Questions: Benefits of Asking More Than One Question. *Public Opinion Quarterly*, 54(3), 330-345.
- Martinez, M. D. (2003). Comment on "Voter Turnout and the National Election Studies". *Political Analysis*, 11(2), 187-192.
- McDonald, M. P. (2003). On the Overreport Bias of the National Election Study Turnout Rate. *Political Analysis*, 11, 180-186.
- Miller, M. (1952). The Waukegan Study of Voter Turnout Prediction. *Public Opinion Quarterly*, 16, 381-398.
- Mummendey, H. D. (1999). *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. 2. Ausgabe. Göttingen/Bern/Toronto/Seattle: Hogrefe.
- Parry, H. J. & Crossley, H. M. (1950). Validity of Responses to Survey Questions. *Public Opinion Quarterly*, 12, 585-598.
- Phillips, D. L. (1973). *Abandoning Method*. San Francisco: Jossey-Bass.
- Presser, S. (1984). Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific? *Public Opinion Quarterly*, 48, 344-355.
- Presser, S. (1990). Can Changes in Context Reduce Vote Overreporting in Surveys? *Public Opinion Quarterly*, 54, 586-593.
- Presser, S. & Traugott, M. (1992). Little White Lies and Social Science Models: Correlated Response Error in a Panel Study of Voting. *Public Opinion Quarterly*, 54, 77-86.
- Presser, S., Traugott, M. W. & Traugott, S. M. (1990). *Vote 'over' Reporting in Surveys: The Records or the Respondents?* National Election Study Technical Report 010157. Michigan.
- Rattinger, H. & Wagner, S. (2003). *Germany: Supplementing Sense of Duty by Cognitive Mobilisation*. Bamberg: University of Bamberg.
- Renzetti, C. M. & Lee, R. (Eds.) (1993). *Researching Sensitive Topics*. Newbury Park: Sage.
- Richman, W. L., Kiesler, S., Weisband, S., Drasgow, F. (1999). A Meta-Analytical Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. *Journal of Applied Social Psychology*, 24(5), 754-775.
- Riklin, A. & Kley, R. (1981). *Stimmabstinenz und direkte Demokratie. Ursachen – Bewertungen – Konsequenzen*. Bern/Stuttgart: Paul Haupt.
- Rogers, T. F. (1976). Interviews by Telephone and in Person: Quality of Responses and Field Performance. *Public Opinion Quarterly*, 40, 51-65.

- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.
- Scholl, A. (1993). *Die Befragung als Kommunikationssituation. Zur Reaktivität im Forschungsinterview*. Opladen: Westdeutscher Verlag.
- Schwarz, N. & Scheuring, B. (1992). Selbstberichtete Verhaltens- und Symptommhäufigkeiten: Was Befragte aus Antwortvorgaben des Fragebogens lernen. *Zeitschrift für Klinische Psychologie*, 21, 197-208.
- Sigelman, L. (1982). The Nonvoting Voter in Voting Research. *American Journal of Political Science*, 26(1), 47-56.
- Silver, B. D. et al. (1986). Who Overreports Voting? *American Political Science Review* 88(2), 613-624.
- Steiner, J. (1965). Überprüfung von Interview-Ergebnissen über die Stimm- und Wahlbeteiligung durch amtliche Angaben. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 17, 234-244.
- Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (Ses-17): Entwicklung und erste Befunde zur Reliabilität und Validität. *Diagnostica*, 45(4), 173-177.
- Sudman, S. & Bradburn, N. B. (1974). *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Swaddle, K. & Heath, A. (1989). Official and Reported Turnout in the British General Election of 1987. *British Journal of Political Science*, 19, 527-541.
- Tourangeau, R., Rips, L. J. & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Traugott, M. (1989). *Validating Self-Reported Vote 1964-1988. Technical Report No. 34 National Election Studies*. Michigan: ISR University of Michigan.
- Traugott, M. W. & Katosh, J. P. (1979). Response Validity in Survey of Voting Behavior. *Public Opinion Quarterly*, 43, 359-377.
- Traugott, S. (1989). *Validating Self-Reported Vote: 1984 - 1988, National Election Study*. Technical Report 010152. Michigan.
- Van der Kolk, H. & Aarts, K. (2001). *Turnout Bias. Non-Response and Stimulus Effects in the Dutch Parliamentary Election Study 1998*. Paper presented at the Annual Meeting of the American Political Science Association, San Francisco.
- Verband Schweizer Marketing- und Sozialforscher (SMS) (2001). *Jahrbuch 2001*. Berneck: SMS.
- Volgy, T. J. & Schwarz, J. E. (1984). Misreporting and Vicarious Political Participation at the Local Level. *Public Opinion Quarterly*, 48, 757-765.
- Voogt, R. (2000). *An Alternative Approach to Correcting Response and Nonresponse Bias in Election Research*. Amsterdam: University of Amsterdam.

- Voogt, R. & Saris, W. E. (2003). To Participate or Not to Participate: The Link between Survey Participation, Electoral Participation, and Political Interest. *Political Analysis*, 11(2), 164-179.
- Vowles, J. et al. (1995). *Towards Consensus: The 1993 Election in New Zealand and the Transition to Proportional Representation*. Auckland: Auckland University Press.
- Wolfinger, R. & Rosenstone, S. J. (1980). *Who Votes?* New Haven: Yale University Press.
- Wright, G. C. (1993). Errors in Measuring Vote Choice in the National Election Studies, 1952-1988. *American Journal of Political Science*, 37, 291-316.

Abkürzungsverzeichnis

AHV	Alters- und Hinterlassenenversicherung
ANES	American National Election Study
GfS	Gesellschaft für praktische Sozialforschung
IHA-GfK	Meinungsforschungsinstitut Schweiz, Hergiswil, Marktleader
IPSO	Meinungsforschungsinstitut Schweiz, Zürich; gehört mittlerweile zu IHA
IV	Invalidentversicherung
JSTOR	The Scholarly Journal Archive
OR	Overreporting
SIDOS	Schweizerischer Informations- und Daten-Archivdienst für die Sozialwissenschaften
SNF	Schweizerischer Nationalfonds
VOX/VOX-Analyse	Nachbefragung zu den eidgenössischen Urnengängen, 1977 ff.

Korrespondenzadressen

Prof. Dr. Sibylle Hardmeier
 Forschungsprofessorin
 Wissenschaftszentrum Berlin für
 Sozialforschung (WZB)
 Reichpietschufer 50
 10785 Berlin
 E-mail: hardmeier@wz-berlin.de

Marie-Christine Fontana, lic. phil.
 Institut d'Etudes Politiques et Internationales
 Bât. Humense
 Université de Lausanne
 1015 Lausanne
 Schweiz
 E-mail: Marie-Christine.Fontana@unil.ch